引用格式： 方思怡. 大小模型协同驱动的标准文献语义匹配方法与应用研究[J]. 标准化学报，2026(11): 17-23.
FANG Siyi. Research on Standard Literature Semantic Matching Method and Application Driven by Large and Small Models Collaboratively[J]. Journal of Standardization, 2026(11): 17-23.

大小模型协同驱动的标准文献语义匹配方法与应用研究

方思怡

（上海市质量和标准化研究院）

摘　要：【目的】文本语义匹配是推动标准数智化深入发展的关键技术，本文提出面向标准数智化转型的标准文献语义匹配框架以支撑标准数智化建设。【方法】结合新兴技术与标准数智化需求，开展了大小模型协同驱动下的标准文献语义匹配方法与应用研究。【结果】构建了融合大模型与轻量级小模型的标准文献语义匹配框架，并以ICS与国民经济行业分类的跨体系映射为例验证了该范式的可行性，形成了ICS与国民经济行业分类在中类层级上的映射数据集。【结论】为涉及标准文献语义匹配的标准数智化应用场景建设提供技术参考，助推标准数智化的深入发展。

关键词： 语义匹配；标准数智化；大模型；ICS；国民经济行业分类

DOI编码： 10.3969/j.issn.2097-857X.2026.11.002

Research on Standard Literature Semantic Matching Method and Application Driven by Large and Small Models Collaboratively

FANG Siyi

(Shanghai Institute of Quality and Standardization)

Abstract: [Objective] Text semantic matching is a key technology driving the in-depth development of standard digitalization and intelligentization. This paper proposes a standard literature semantic matching framework for the transformation towards standard digitalization and intelligentization. [Methods] Combining emerging technologies with the needs of standard digitalization and intelligentization, research on standard literature semantic matching methods and applications driven by the collaboration of large and small models has been conducted. [Results] A standard literature semantic matching framework integrating large models and lightweight small models has been developed, and the feasibility of this paradigm has been verified through an example of cross-system mapping between the ICS and industrial classification for national economic activities, resulting in a mapping dataset of the ICS and industrial classification for national economic activities at the medium category level. [Conclusion] This provides technical reference for the construction of standard digitalization and intelligentization application scenarios involving standard literature semantic matching, promoting the in-depth development of standard digitalization and intelligentization.

Keywords: semantic matching; standard digitalization and intelligentization; large language model; ICS; industrial classification for national economic activities

---

基金项目： 本文受上海市质量和标准化研究院院立项目“生成式人工智能视角下的标准专题库建设及应用研究”（项目编号：YRY202610）资助。

作者简介： 方思怡，硕士，工程师、知识产权师，研究方向为标准数智化、标准知识服务。

0 引言

作为自然语言处理（Natural Language Processing, NLP）领域的关键基础性任务，文本语义匹配（Text Semantic Matching, TSM）旨在通过特定算法衡量不同语义片段之间的语义相似度[1]，是支撑问答系统、信息检索、机器翻译、文本理解等多个自然语言处理应用场景的重要基石[2]。随着标准数字化和智能化转型的深入推进，文本语义匹配也将逐渐在标准文献业务中发挥关键作用，其核心价值主要体现在两方面：一是通过建立标准文献内容及衍生信息之间知识关联的方式来推动数字标准馆的标准数智底座建设，为构建标准领域高质量数据集提供重要基础；二是作为核心技术来支撑涉及语义匹配的标准数智化应用场景建设，应用场景包括但不限于标准智能检索、标准智能推荐、标准智能问答、标准体系与专题库智能搭建、标准智能伴读、标准智能查新等。

本文结合新兴技术发展趋势，根据标准数智化应用场景的建设需求，提出了融合大语言模型（Large Language Model, LLM）与轻量级小模型的标准文献语义匹配框架，并以标准文献跨体系分类匹配为应用对象，实现了国际标准分类法（ICS）分类信息与国民经济行业分类信息在特定层级的映射，并在实践层面验证了该框架的可行性，为标准数智化转型所需的数智资源建设提供了一定的技术参考。

1 国内外标准相关领域的语义匹配研究现状分析

标准文献是在标准化活动中根据既定规则，经各相关利益方协商一致后形成的技术性文件[3]。从性质上来看，它既是科技情报资源的一大分支[4]，也是国家质量基础设施（National Quality Infrastructure, NQI）的重要组成部分。文本语义匹配是一项广泛应用于多个领域的自然语言处理技术，已历经多个发展阶段。与其他领域类似，以图情为代表的标准相关领域中的文本语义匹配技术也包括基于传统机器学习、基于神经网络模型、基于预训练模型、基于知识增强这4个类型[5]。基于传统机器学习的文本语义匹配技术主要从字符串、词频-逆文档频率（Term Frequency-Inverse Document Frequency, TF-IDF）和主题模型这3个方面出发[6-7]计算不同文本的字面重叠程度，其优势在于操作简便，缺点是无法揭示文本的深层语义关联，通常可作为数据挖掘与知识组织中间环节的技术支撑。例如在标准知识图谱构建流程中采用余弦相似度来确定标准文献元素的语义关系[8]。基于神经网络模型的文本语义匹配技术主要有浅层神经网络模型、循环神经网络模型、卷积神经网络模型等，其中浅层神经网络模型以Word2Vec[9]、FastText[10]等静态嵌入模型（Embedding Model）为主。相较于浅层神经网络模型，循环神经网络与卷积神经网络模型主要通过文本序列编码的方式捕捉文本序列的局部或全局特征，在语义理解上更为出色，对特征工程的依赖也更低。与上述传统深度学习模型相比，预训练模型（Pre-trained Models, PTMs）经由大规模文本语料训练而来，更擅长自动理解文本的深层语义信息，包括BERT及其一系列优化和改进后的变式模型，如ALBERT、KBERT、RoBERTa[11]、SpanBERT、SBERT、FastBERT等。为了解决长久以来文本语义匹配中存在的关键信息捕捉不足[12]、上下文欠缺和语义混淆等问题，基于知识增强的范式应运而生，旨在通过引入外部知识源的方式提高语义匹配的准确性，包括XLNet、ERNIE[13]、KIM、CEUR等。

近年来，随着生成式人工智能技术的蓬勃发展，文本语义匹配技术也步入新的发展阶段。大模型凭借其优异的内容生成、语义理解、知识整合、逻辑推理等通用能力[14]成为提升文本语义匹配任务准确度的关键。与传统的机器学习和深度学习模型相比，大模型在跨领域知识迁移上表现更佳，但也普遍存在上下文长度限制[15]、AI幻觉[16]等不足。为了解决上述问题，大小模型协同范式逐渐成为自然语言处理技术的发展新趋势。与大模

型相对，小模型（Small Model，SM）是指参数规模小、部署轻便、结构精简的模型，在广义上泛指参数量较少的模型，既包括小型语言模型（Small Language Model，SLM），也包括简单的统计模型、浅层神经网络模型等[17]，其典型代表为基于传统机器学习、深度学习、数理统计等构建而成的数据驱动模型[18]。当前，大小模型协同范式的有效性已在实体抽取[19-20]、知识图谱构建[21]等方面得到一定的验证，但在文本语义匹配中的技术方案还有待细化。标准领域的数智化转型尚处于起步阶段，亟须在大小模型协同视角下，结合标准文献业务的实际需求，形成适用于标准文献的文本语义匹配范式。

2 研究方法与结果

2.1 大小模型协同驱动的标准文献语义匹配策略

2.1.1 大小模型协同驱动的标准文献语义匹配框架

为了能够充分利用大模型与广义范畴下的小模型在文本语义匹配中的优势，本文结合标准文献的数据特性与应用需求，提出了融合大模型与轻量级小模型的标准文献语义匹配范式，设计了大小模型协同驱动的标准文献语义匹配的整体框架（详见图1）。

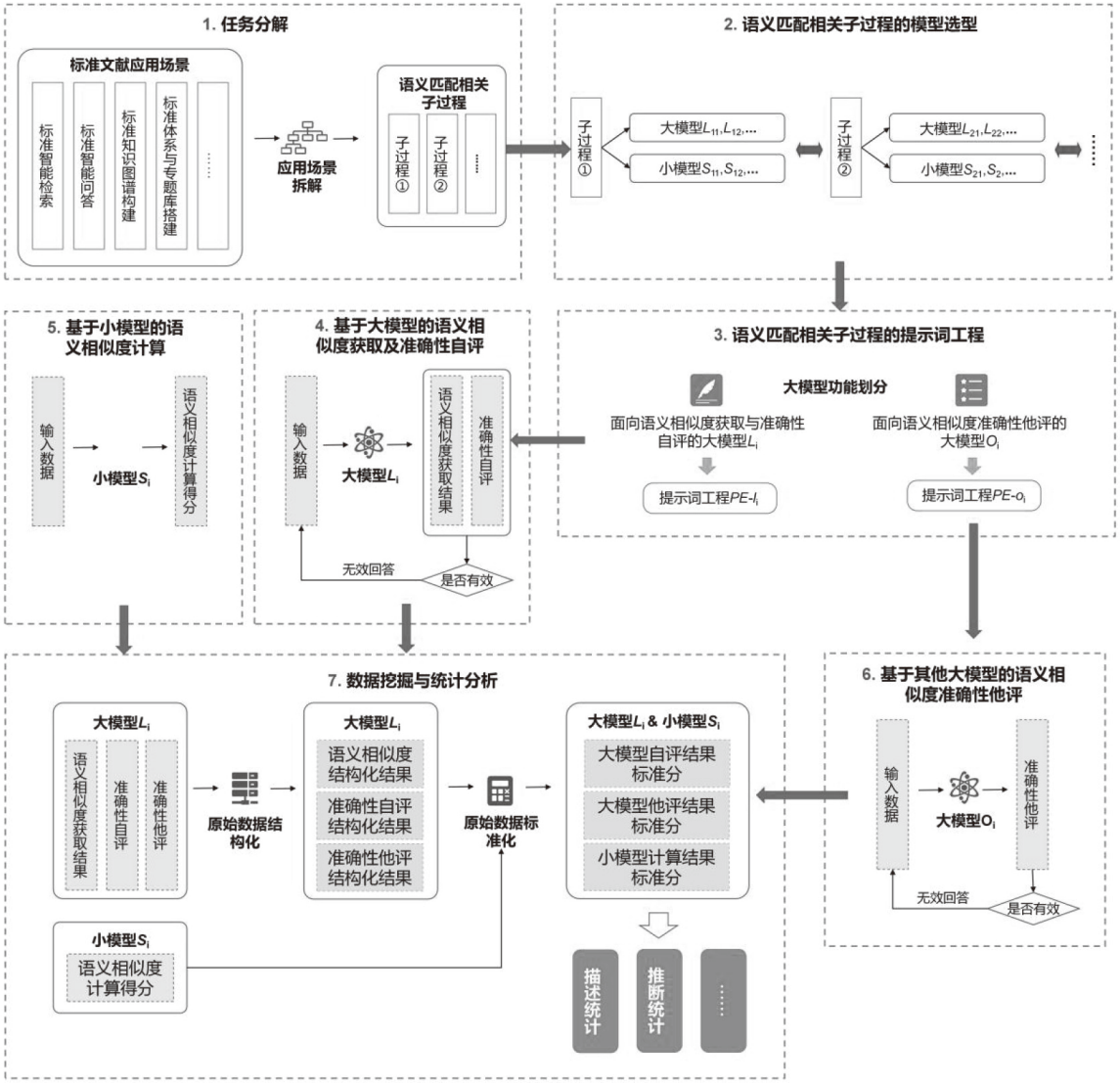


图1 大小模型协同驱动的标准文献语义匹配框架

大小模型协同驱动的标准文献语义匹配框架主要包括任务分解、模型选型、提示词工程等7个核心流程。与以往研究相比，本文提出的语义匹配框架有以下2个特点：

（1）设置大模型准确性评价环节。为了能够评估大模型回答的准确性，本文在标准文献语义匹配框架中引入了基于李克特量表（Likert Scale）的大模型准确性自评与他评机制。其中，准确性自评是指某个大模型根据自身回答做出的准确性评分，准确性他评是指其他大模型对该大模型回答做出的准确性评分。近年来多项研究显示，基于大模型的评分与人工评分具有一定的相似性和参考价值[22-23]。因此，本文将大模型评分纳入框架，形成多个大模型互相验证的模式，以此提升大模型回答的准确性。

（2）提出基于标准分数的数据分析思路。本文提出的标准文献语义匹配框架可包含若干种大模型和小模型。为了能够对不同模型的结果进行有效融合和科学比较，将大模型和小模型的运行结果统一转化为标准分数并基于此进行统计分析。

2.1.2 大小模型协同驱动的标准文献语义匹配流程

上述7个核心流程的具体说明如下：

在任务分解阶段，需要综合技术、业务等角度对标准文献应用场景进行任务拆解，划分为若干子过程，并明确与标准文献语义匹配密切相关的子过程。

在本文提出的框架中主要有2类大模型：一是获取语义相似度结果并进行准确性自评的大模型L_i，二是根据大模型L_i的语义相似度结果进行准确性他评的大模型O_i。在模型选型阶段，主要针对语义匹配相关的所有子过程分别进行大模型与小模型的选型测试，通常包括需求分析、备选模型列表制定、模型测试与评估等步骤。在提示词工程阶段，根据大模型的功能划分，对框架中不同功能的大模型分别确定与功能匹配的提示词。

在大模型的语义相似度获取与准确性自评阶段，根据子过程的具体需求，获取大模型Li的语义相似度结果及准确性自评得分，并对大模型Li回答的有效性进行检验，对于无效回答需重新获取上述结果，直至大模型L_i的回答通过有效性检验。在小模型的语义相似度计算阶段，采用模型选型阶段确定的小模型开展语义相似度计算。与大模型L_i自评流程类似，在其他大模型的语义相似度准确性他评阶段，采用其他大模型O_i对大模型O_i的回答进行准确性他评并进行有效性检验，对无效回答需重新获取上述结果，直至大模型O_i的回答通过有效性检验。

在数据挖掘与统计分析阶段，首先对大模型的运行结果进行结构化处理，再将大模型的结构化结果与小模型的运行结果统一转化为标准分数，并根据实际的业务需求对大、小模型的标准分数开展统计分析。

2.2 大小模型协同驱动的标准文献语义匹配应用

2.2.1 任务说明

文本分类是自然语言处理中的一项基本任务，以文本语义匹配为关键环节，旨在根据文本的语义内容实现类别划分[24]。标准文献分类是遵循一定的体系揭示标准文献内容的有效方式[25]，对实现标准文献资源科学管理和开展标准知识服务具有重要的作用。从标准文献分类的历史来看，我国自1997年起在国家标准、行业标准和地方标准这3个类别中采用《国际标准分类法》（以下简称ICS）与《中国标准文献分类法》（以下简称CCS）并行的分类模式[26-27]，两者均以产业划分为依据。与CCS相比，ICS在国际上的适用性更为广泛，但与我国的经济活动关联性不够强。针对全社会经济活动的分类需求，我国发布了GB/T 4754—2017《国民经济行业分类》[28]，该标准以经济活动性质为依据进行行业划分，能够准确而及时地反映我国经济发展和产业结构转型升级过程中涌现的新产业[29]，更契合我国的国情需要，对行业高质量发展至关重要。为了能够兼具国际和国内的产业发展适用性，进一步丰富标准文献分类体系，本文采用大小模型协同驱动的标准文献语义匹配框架，根据ICS与国民经济行业分类的数据特点，建立ICS与国民经济行业分类的映射关系。考虑到国民经济行业分类具有门类、大类、中类、小类

4个层次，ICS则包含一级、二级和三级类目，两者在分类层级上存在差异。本文聚焦ICS与国民经济行业中类层次的语义匹配，实现ICS所有类目与国民经济行业“门类—大类—中类”的知识关联。

2.2.2 数据概况

本文的ICS分类数据集的原始数据集来自上海标准文献馆的馆藏数据资源，国民经济行业分类数据集的原始数据来自GB/T 4754—2017及其修订信息。所有原始数据均被规范化处理，形成相应的结构化数据形式。

2.2.3 主要流程

本文根据任务实际需求，在大小模型协同驱动标准文献语义匹配的框架基础上建立ICS与国民经济行业中类的映射关系，具体流程如下：

（1）在任务分解阶段，为了能够兼顾算力资源与映射效率，本文将任务拆解为以下2个步骤：一是根据ICS分类名称获取与之语义匹配的门类名称；二是以该门类下属的所有中类为范围，继续获取与ICS分类名称语义匹配的中类名称。

（2）在模型选型阶段，对ICS分类名称与国民经济行业分类名称进行文本特征分析。两者均为长度相对较短的中文字符串，且涉及多个行业，因此在大模型方面宜选取跨领域能力较强的通用大模型。在小模型方面，已有研究将文本语义相似度算法划分为统计算法、神经网络、基于语料库/知识库的测量方法[30]。本文以便捷性和实用性为优先条件，在广义范畴的小模型内选取契合任务需求的简单统计模型，构建了融合多种基于字符串的语义相似度算法的混合模型。经过一系列选型测试，本文最终确定了大小模型的型号，如表1所示。

（3）在大模型的提示词工程阶段，本文设计了适用于不同任务子过程的大模型提示词。在基于大模型的语义相似度结果获取与准确性自评阶段，根据任务分解，首先采用Spark 4.0 Ultra获取与ICS分类名称语义相似度较高的门类名称并进行准确性自评，在确保所有回答均有效且通过其他大模型的语义准确性他评后，自动提取ICS分类名称对应门类下的所有中类，采用SparkMax获取与ICS分类名称语义相似度较高的中类名称并进行准确性自评。

（4）在基于小模型的语义相似度计算阶段，考虑到门类名称通常为较为广泛的分类信息，与ICS分类名称的字面相似度普遍不高，而中类名称在字面上更可能与ICS分类名称存在重叠性，更适用于基于字符串的语义相似度算法。因此，本文主要采用融合编辑距离、杰卡德系数和余弦相似性算法的模型来计算ICS分类名称与大模型获取中类名称的语义相似度，不对门类名称进行上述方法的语义相似度计算。

（5）在基于其他大模型的语义相似度准确性他评阶段，由于本文的任务涉及2次大模型的语义相似度结果获取，需要分别采用ERNIE-Speed-128k、DeepSeek-V3.2对ICS分类名称及对应门类和中类的语义相似度进行准确性他评。

（6）在数据挖掘与统计分析阶段，由于ICS分类名称对应的中类名称是本文任务的最终目标，因此该阶段的核心环节是将大模型对中类获取结果的准确性评价得分与小模型的语义相似度计算结

表1 ICS与国民经济行业分类语义匹配的大小模型选型结果

框架中的模型分类	模型功能	模型名称
大模型	获取与ICS分类名称语义匹配的门类名称以及准确性自评得分	Spark 4.0 Ultra
	在大类下的中类范围内，获取与ICS分类名称语义匹配的中类名称以及准确性自评得分	SparkMax
	分别对ICS分类名称与门类名称、中类名称的语义相似度进行准确性他评	ERNIE-Speed-128k、DeepSeek-V3.2
小模型	对ICS分类名称与中类名称进行语义相似度计算	基于编辑距离、杰卡德系数、余弦相似性算法的混合模型

果统一转化为*Z*分数，从而降低不同模型在评价尺度上的差异性，并进一步对大模型准确性评价得分和小模型计算结果进行假设检验，根据假设检验的统计结果明确大模型自评与他评的合理性，最后计算ICS分类名称与其对应所有中类名称的大小模型运行结果的*Z*分数总和，选取*Z*分数总和最大值作为该ICS分类的中类名称最优解，并自动提取该中类名称所属的“门类—大类—中类”名称及代码组合。

2.2.4 应用结果

对大模型获取的ICS分类名称对应门类名称的准确性自评与他评*Z*分数进行统计分析，结果显示，ICS分类名称与对应门类的Spark 4.0 Ultra准确性自评Z分数与ERNIE-Speed-128k、DeepSeek-V3.2准确性他评*Z*分数均存在显著相关（$P<0.01$），对上述三者分别进行两两之间的配对样本*t*检验，发现均不存在显著差异（$P>0.05$）。该结果在一定程度上说明了本文选取的大模型在门类名称的语义相关性判断上具有较高的一致性。同理，本文对ICS分类名称与对应中类的SparkMax准确性自评*Z*分数与ERNIE-Speed-128k、DeepSeek-V3.2准确性他评*Z*分数也进行了两两之间的相关分析与配对样本T检验，结果同样显示上述三者的*Z*分数之间均存在显著相关（$P<0.01$），也均不存在显著差异（$P>0.05$）。该结果表明本文所选取的大模型在中类名称的回答上具有高度一致性，在一定程度上印证了ICS分类名称与对应中类的语义匹配结果的可靠性。ICS分类名称与国民经济行业分类的部分映射结果如表2所示。

3　结 语

本文从标准数智化转型的实际需求出发，结合新兴技术发展趋势，提出了融合大模型与轻量级小模型的标准文献语义匹配方法，通过采用大小模型协同驱动的标准文献语义匹配框架，实现了ICS分类名称与国民经济行业分类在中类层级上的映射，形成了ICS分类名称与国民经济行业“门类—大类—中类”名称及代码的映射数据集，在一定程度上建立了当前国内主流的标准分类法与我国经济活动的关联性。后续，将从以下2方面开展研究工作：一是持续优化标准文献语义匹配框架。后续研究将在框架中融入人工匹配与审核环节，从而实现大模型、小模型与人工“三位一体”的标准文献语义匹配机制。二是深入开展标准数智化应用场景建设。标准文献语义匹配是开展标准数智化应用的基础性技术。本文采用标准文献语义匹配方法，初步实现了ICS与国民经济行业分类在中类层级上的跨体系匹配，后续将考虑在标准智能检索、标准智能问答、标准智能推荐、标准体系与专题库智能搭建、标准智能查新等标准业务场景中实现ICS与国民经济行业分类映射数据集的落地应用，并进一步采用该框架完成标准核心要素与衍生信息的标准知识语义网络构建，建设标准领域的行业高质量数据集，助力数字标准馆落地，从而推动标准数智化转型走深走实。

表2 ICS分类名称与国民经济行业分类的部分映射结果

ICS分类号	ICS分类名称	对应的“门类—大类—中类”		Z分数总和
		名称	代码	
11.060.20	牙科设备	制造业—专用设备制造业—医疗仪器设备及器械制造	C-35-358	4.25
11.060.25	牙科仪器	制造业—专用设备制造业—医疗仪器设备及器械制造	C-35-358	4.82
11.080	消毒和灭菌	卫生和社会工作—卫生—专业公共卫生服务	Q-84-843	1.70
11.080.01	消毒和灭菌综合	卫生和社会工作—卫生—专业公共卫生服务	Q-84-843	0.28
11.080.10	消毒设备	制造业—专用设备制造业—医疗仪器设备及器械制造	C-35-358	4.33
11.080.20	消毒剂和防腐剂	制造业—化学原料和化学制品制造业—日用化学产品制造	C-26-268	1.59

参考文献

[1] 杨瑜. 基于深度学习的文本语义匹配研究[D].贵阳：贵州大学，2024.

[2] 李壮壮.基于预训练模型的中文文本语义匹配方法研究[D]. 济南：山东师范大学，2024.

[3] GB/T 1.1—2020 标准化工作导则 第1部分：标准化文件的结构和起草规则[S].

[4] 赵伟，张览，望俊成. 标准文献知识图谱构建的模型设计与集成方法[J]. 情报工程，2021，7（6）：58–66.

[5] 马浩洋. 面向中文文本的语义匹配模型研究与实现[D]. 北京：中国电子科技集团公司电子科学研究院，2024.

[6] 王宝军. 中文文本数据相似度匹配算法及性能分析[J]. 电子制作，2025，33（12）：50–53.

[7] Kim D，Seo D，Cho S，et al. Multi–co–training for document classification using various document representations： TF–IDF，LDA，and Doc2Vec[J]. Information Sciences，2019，477：15–29.

[8] 林杰，周育忠，宋禹飞，等. 大模型在标准数据结构语义化中的技术应用与实现[J]. 科技与创新，2025（8）：222–224，228.

[9] 丁敬达，陈一帆，刘超，等. 基于共词和Word2Vec加权向量的文献–主题语义匹配分析方法[J]. 图书情报工作，2022，66（12）：108–116.

[10] 王树刚，郭涛，朱超，等. 基于FastText模型和LCS匹配算法的虚端子自动关联方法[J]. 电力信息与通信技术，2025，23（4）：70–76.

[11] 向尕，胡演，张仰森，等. 面向威胁情报分析的恶意软件知识图谱构建[J]. 应用科学学报，2026，44（1）：67–82.

[12] 张一鸣，朱子杰，张霜，等. 基于动态词遮掩和知识增强的语义匹配模型[J]. 中国电子科学研究院学报，2025，20（5）：520–528.

[13] 王钦晨. 基于深度学习的短文本语义匹配策略研究[D]. 太原：太原理工大学，2024.

[14] 李昀峰，赵泓辰. 大模型驱动下的汽车标准数字化发展路径分析[J]. 汽车与配件，2026，46（6）：62–65.

[15] 魏蕴田，金苍宏，费峥东，等. 大小模型协同优化的兴趣点轨迹预测框架[J]. 浙江大学学报（工学版），2026，60（6）：1251–1260.

[16] 叶菲楠，李侠. 预测加工视角下的AI幻觉：本质、机制与治理[J]. 科学与管理，2026，46（5）：23–29.

[17] Chen L， Varoquaux G. What is the role of small models in the LLM era： a survey[A]. arXiv， 2024. DOI：10.48550/arXiv.2409.06857.

[18] 张超，袁梓涵，周光辉，等. 大小模型协同的工业AI Agent：概念、框架与应用[J]. 中国科学：技术科学，2026，56(4)：1072–1084.

[19] 梁柱，刘寅鹏，石湘，等. 面向科技实体抽取任务的大小模型少样本数据协同训练研究[J]. 信息资源管理学报，2025，15（4）：129–143.

[20] 冯国平，陈志坚，林志煜，等. 基于动态领域图谱与大小模型协同的电力领域术语识别[J/OL]. 计算机工程，2026–04–23. https：//doi.org/10.19678/j.issn.1000–3428.0252291.

[21] 邸晓丽. 大小模型协同的知识图谱补全方法研究[D]. 太原：太原理工大学，2025.

[22] Gu J，Liu J，Zeng L，et al. Comparing ChatGPT and validated questionnaires in assessing loneliness and online social support among college students： a cross–sectional study[J]. Scientific Reports， 2025，15（1）：20621.

[23] Lu Y，Liles X，Ma X. GenAI and human assessments of L2 Chinese writing：Interrater reliability and rater bias[J]. Assessing Writing，2025,66：100989.

[24] 袁辉，赵捷，侯博，等. 基于深度学习的法人和其他组织国民经济行业分类标准化流程研究[J]. 中国标准化，2024（5）：61–65.

[25] 许莉莉，牟长青，汤海荣. 海洋标准文献分类的理论与实践[J]. 中国标准化，2017（12）：70–71.

[26] 汪海波，汤彩霞. 标准文献CNMARC格式分类标引的再探讨[J]. 图书馆杂志，2026，45（2）：65–68.

[27] 程军毅. 《中国标准文献分类法》与《国际标准分类法》在我国标准文献中的应用[J]. 中国标准导报，2014（5）：46–49.

[28] GB/T 4754—2017 中国标准化研究院.国民经济行业分类[S].

[29] 国家统计局. 国民经济行业分类包括哪些类别[EB/OL]. （2025–03–21）[2026–04–23]. https：//www.stats.gov.cn/zs/tjws/tjbz/202301/t20230101_1903769.html.

[30] Singh A，Garg S K. Comparative study of different document similarity measures and models[J]. Assessing Writing，2024，66：737–746.